Towards Automatic Modeling for Cultural Heritage Application

TOWARDS AUTOMATIC MODELING FOR CULTURAL HERITAGE APPLICATIONS

A. Guarnieri[1] G. Guidi[2] G. Tucci[3] A. Vettore[1]

[1] Cirgeo – Interdept. Research Center of Geomatic, University of Padua, Italy – cirgeo@unipd.it
[2] Dept. Of Electronics and Telecomm., University of Florence, Italy - g.guidi@ieee.org
[3] Dept. of Scienze e Tecniche per i Processi di Insediamento, Polytechnic of Turin, Italy – tucci@archi.polito.it

KEYWORDS: 3D modeling, range data alignement, laser scanning, cultural heritage documentation, spin image, virtual restoration

ABSTRACT

Recent advances in the field of laser scanning technology along with the availability of more powerful computing resources have favoured the increasing interest of surveyors, architects, archaeologists towards laser scanners as a very promising alternative for cultural heritage surveying. Thousands of points can be acquired in a few seconds with an accuracy that is adequate to build 3D models for single objects so as for whole environments. At the present, resulting 3D digital models offer an invaluable mean for documentation, archiving, structural analysis and restoration of the large amount of objects belonging to our historical and cultural heritage. Usually, the end products of the whole workflow (survey and modeling) are VR representations (Vrml, Flash), movies (AVI, DVx, Mpeg), Digital Surface Models (DSM) and orthophotos as well. The creation of a 3D model requires a lot of data about the object surface or volume, which have then to be aggregated, regardless the data format and the acquisition device used. In most cases, the data registration step is based on ICP, that iterativelly finds the mutual orientation between two range maps, starting from an initial guess given by an operator. This approach is often time-consuming, increases the final cost of the 3D model and represents the major limit to the wide spreading of real object models.

In this paper an overview of our automatic range data registration system is presented, focusing on the integration between the two main blocks. In the first one, overlapping areas between range image pairs are detected by mean of spin-images and an initial approximate alignement between image pairs is computed. Then, in the second block, a refinement of this estimate is performed by use of a cascade of two registration algorithms: the Frequency Domain and the ICP.

1. INTRODUCTION

We are well aware that nowadays the information plays a major role in our life. This concept affects greatly the work of the surveyor too. As time goes on photogrammetrist and surveyor are requested to provide a complete and global information about the surveyed object regardless its size and shape complexity. Whether we deal with close-range or long-range surveys, whether the target is a single and simple object or a more complex and structured one or a piece of land, both metric and qualitative information are requested in the final product. This approach has favoured in last years the rise of the interest towards the laser scanning technology, as a very promising alternative for surveying applications. Airborne and ground-based laser scanners allow to acquire very quickly a huge amount of 3D data which can be often profitably combined with color high-resolution digital images to provide a 3D representation of the environment where we live. A major advantage of 3D models relies on the fact that they allow to represent real objects more adequately than through a single picture or collection of pictures, by providing a higher level of detail together with a good metric accuracy. These models are currently used for cultural heritage, industrial, land management or also medical applications. In the cultural heritage field, 3D models represent an interesting tool for as-built documentation and interactive visualization purposes, e.g. to create virtual reality environments. In some cases (El-Hakim, 2001) 3D models obtained by laser scanning were used to fill a virtual environment with real objects, in order to get a faithful copy of a real environment, such as the interior of a museum or historical building. Furthermore, in these VR applications the user point of view can be easily changed, providing in this way a useful tool by which the human direct inspection can be well simulated. Beyond VR applications, laser scanning data can be also used to generate orthophotos in relative short time, though sometime the quality of the final product is not at the same level of the photogrammetric approach. From metric point of view, attempts are being undertaken in order to extract as more automatically as possible from the 3D model the main vector lines defining the shape of the object. Indeed, a vectorized representation is still more appreciated by end-users as it can be more easily processed in CAD environments.

As mentioned before, 3D models from range data are composed by thousands or even million of points. They can be then well managed only if a suitable hardware platform is available. Main requirements in this sense are powerful CPUs and a large amount of system memory (512 MB of Ram or higher). In last years different methods were proposed in order to create 3D models even with limited hardware resources (Guarnieri,1997), but continuous advances in computer science field have already made out-of-date such solutions. The rapid succession on the marketplace of new and more powerful hardware components has greatly favoured the use of laser scanners on one hand, but it is also changing the approach towards the way surveys with laser scanner are performed, on the other hand. Nowadays creating and managing 3D models made of million of points is mainly limited by not enough powerful hardware resources. However this is only a technological (and therefore temporary) limit, not a physical one. Getting more processing power is just a matter of time, often of a few months only.

It is then advisable that, regardless of the kind of target to be surveyed, the more profitable approach consists to acquire once the range data at high resolution, maximizing as much

as possible the level of achievable information. Clearly, if not supported by present hardware, a low resolution 3D model can be provided as first product. Anyway the full resolution one will be created in the following, as soon as suitable computing resources will be available.
It seems therefore that laser scanning technology, along with digital imaging, could satisfy in next future the requirement of a multipurpose complete information, better than other approaches.

Despite the advantages of laser scanning-based survey, new set of issues has to be addressed in the 3D modeling of real objects. Actually, 3D modeling of real free-form surfaces consist of the following steps:

1. Manual pairwise alignment of the 3D images;
2. Global alignment;
3. Fusion of the 3D data originally captured as clouds of points into 3D surfaces;
4. Editing of possible surface holes due to minor missing data.

Step 2) and 3) are already performed automatically, while step 4) may not be necessary if an adequate amount of data is captured (which however may not always be feasible).
At the present, step 1) represents the most time-consuming issue in 3D modeling, overall in the field of cultural heritage applications, where object modeling is affected by a number of difficulties such as: the shape, typically more articulated than that of mechanical objects; the size, which may not be small; the fact that objects cannot be taken into a laboratory but almost always need portable equipment; the required precision, if physical duplication has to be included among the possible model's uses. The currently adopted registration techniques can be summarized in two main classes: a feature based approach and a combined laser scanning/classical topographic-photogrammetric approach. In the first one, common features between a range data pair are manually detected by a human operator, or automatically selected by means of primitive fitting, in order to provide a first rough alignment, which is then refined through step2). In the second case, a set of Ground Control Points (GCP), measured with classical topographic-photogrammetric instruments, are employed as constraints in the range data alignment (Guidi et al., 2003). Anyway, both these methods are time-consuming and increase the final cost of the 3D model. The lack of the automatization for 3D data alignment still represents the major limit to the wide spreading of real object: solutions for this issue will be a major progress for general 3D modeling.
In the light of topics previously exposed, an automatic range data registration system has been developed, which is able to execute all the steps needed for 3D modeling of real objects minimizing as more as possible the human intervention, without any other information but the range data only. The work drew the idea from A. E. Johnson, which proposed an innovative solution for the recognition of similarities between 3D surfaces, introducing the *spin-image* concept (Johnson, 1997). The advantage of this approach relies on high computational robustness and effectiveness, which allows to employ standard market-level CPUs. On the ground of the spin-image concept, a full data registration system was developed, in which firstly overlapping areas between two adjacent data sets are automatically detected and then the whole data sets are aligned each other. The registration system is composed by two main blocks. In the first one, overlapping areas between range image pairs are detected by mean of spin-images and an initial approximate alignment between image pairs is computed. Then, in the second block, a refinement of this estimate is performed by using a cascade of two registration algorithms: the Frequency Domain technique (Lucchese, 1997) and the ICP (Besl, 1992).
So far, this system has been applied to register range data of objects of limited size, such little statues or bas-relieves. Usually each data set was composed by a grid of 256x256 points, acquired by a close range laser scanner, which were succesfully pairly aligned in relative short time (3-4 min. each pair). Actually our effort is aimed to improve the system's performance in order to register automatically 3D images of more complex objects, like ancient buildings, churches, bridges and any other infrastructure of cultural and historical interest.
The paper is structured as follows. Section 2 provides the necessary background on the subsytem developed for the rough transformation estimate between a view pair, while section 3 focuses on the framework of the second block, i.e. the subsystem for the registration refinement. Then in section 4 some remarks are discussed on future developments and finally section 5 reports the conclusions.

2. SUBSYSTEM FOR COMMON AREA DETECTION

Though the topic discussed in this section has been already presented in more detailed way (Bologna, 2002), we will provide to the reader a brief overview of the features of the subsystem for the automatic detection of common areas between image pairs. Given a whole set of range data, acquired for instance by a laser scanner, the registration procedure is applied to each pair of range data. In the first stage, a view pair (e.g. A and B) is selected from the starting set and triangular meshes are built for both 3D views, as basis for the computation of spin-images. Essentialy a spin image of a 3D surface is the recording on a 2D accumulator of the coordinates of all the points of a 3D surface. By way of this technique one may associate a collection of images to a 3D surface mesh, as every point of the surface can generate a spin-image. Two surfaces representing the same object from different view-points will be associated to two sets of different spin images: corresponding points in the common region between two 3D images will have similar spin-images (not identical, due to noise and discretization effects) because spin-images exclusively depend on shape's characteristics. Therefore, with this approach the problem of determining the common region between two 3D images can be turned into the recognition of the most similar images of two image sets, a well studied problem for which a number of techniques are available.
In the currently adopted procedure, two different lists (SpinList) are generated: one comprising the spin-images for all surface points of mesh A, while the other containing only the spin-images of a subset of surface points of the second mesh, B. The degree of similarity between each spin-images of that subset and all the images of mesh A is evaluated by computing and ranking a similarity measure C. Repeating this procedure for each point of subset-mesh B, yields a list of possible correspondences between points of the two meshes. The number and the way the point subset of mesh B are selected, play a prominent role for the global effectiveness and performance of the algorithm. The better approach would involve the selection of a limited number of points belonging to the overlapping area between the meshes. Since no a priori information is available about this zone, the points are choosen randomly.

As the implemented algorithm can find multiple and incorrect correspondences, the evaluation of all combinations of three point correspondences for estimating a plausible view transformation would lead to a combinatoric explosion. Thus, in the second stage of implemented procedure, the number of correspondences candidates is reduced by application of filtering and grouping strategies. Basically, likely candidates to matching are detected firstly by thresholding the similarity measures, then remaining correspondences are grouped together according to a geometrical consistency criterion.
Next, for each group the rototranslation parameters are computed through a least square estimate, based on Horn method (Horn, 1987). Once the two meshes are registered through these parameters, for each point of mesh A the closest point of mesh B is determined. If the distance is less than 1.5 times the mesh resolution, then selected point is most likely belonging to the common area. This criterion allows to easily compute an overlapping zone for each point of the correspondence groups. The optimal transformation is associated with the resulting wider common area.
Finally, the detection of common area between a range image pair is performed in the third stage of the subsystem. Here the point clouds represented by the original, not meshed, range data of the two views are registered each other, using the previously estimated transformation. Conceptually, one can expects that the applied rototranslation would put points on the common area between the two views, very close each other. Again, this zone can be identified as the one composed by the set of corresponding points whose euclidean distance is less than a certain threshold. A point on view A, which don't belongs to the overlapping area, will have its closer corresponding point on view B at a distance greater than selected threshold. The framework of presented subsystem is depicted in figure 1.

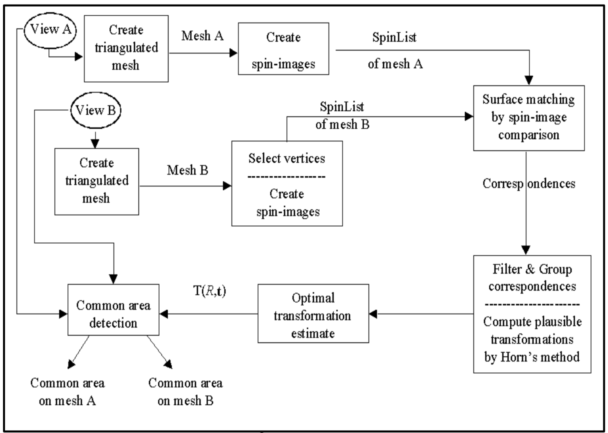

Figure 1: Framework of the overlapping area detection subsystem

3. SUBSYSTEM FOR GLOBAL REGISTRATION

As briefly exposed in previous section, the first block of our registration system provides both the overlapping area and an approximate estimate of the transformation between a 3D view pair, as result of application of Horn's method. Anyway, this estimate cannot be directly used to register the views each other, because of its low degree of accuracy. Therefore a second block has been introduced in the registration process, as depicted in figure 2. That block works as subsystem for the refinement of previous computed alignement, and comprises a cascade of two global registration algorithms: the Frequency Domain and the well known ICP.
Though the first one has been already discussed (Lucchese, 1997), we will recall here a brief overview of its properties. The Frequency Domain tecnique represents an alternative procedure for 3D motion estimate, based on the Fourier transform of the 3-D intensity function, implicitly described by the registered time-sequences of range data. It is a global approach because it operates in the frequency domain using the whole image information and not just a selected subset of the image as the feature-based methods do. Basically, the main advantage of this technique relies on the fact that in the frequency domain rotation and translation can be decoupled, therefore they can be estimated separately.
Indeed, let be $l_1(\mathbf{x})$, $\mathbf{x} \in R^2$ and $l_2(\mathbf{x}) = l_1(R^{-1}\mathbf{x}-\mathbf{t})$, $\mathbf{t} \in R^3$, respectively the common area detected on mesh A and the corresponding rototranslated version on mesh B; then denoting with $L_i(\mathbf{k}) = F[l_i(\mathbf{x})|\ \mathbf{k}]$, $\mathbf{k} = [k_x, k_y, k_z]$, the 3-D cartesian Fourier transform of $l_i(\mathbf{x})$, i = 1, 2, we get:

$$L_i(\mathbf{K}) = F\left[l_i(x) \,|\, \hat{\mathbf{E}}\right] = \iiint_{-\infty}^{+\infty} l_i(x) e^{-j2pK^T x} dx \quad (1)$$

where $\mathbf{K} = [k_x, k_y, k_z]^T$

According to these definitions, it is easy to demonstrate the two Fourier transforms relates as follows:

$$L_2(\hat{\mathbf{E}}) = L_1(R^{-1}\mathbf{K}) e^{-j2pK^T Rt} \quad (2)$$

From (2) one sees that the translation **t** affects only phases and not magnitudes. Magnitudes are related as

$$\left|L_2(\mathbf{k})\right| = \left|L_1(R^{-1}\mathbf{k})\right| \quad (3)$$

and (3) can be used in order to determine R. Therefore in the frequency domain the estimation of R and **t** can be decoupled and one can estimate first R from (3) and then **t** from inverse Fourier transform of phase correlation btw. L_1 and L_2 (4).

$$Q(\mathbf{k}) = \frac{L_1^*(\mathbf{k}) L_2(\mathbf{k})}{\left|L_1(\mathbf{k}) L_2(\mathbf{k})\right|} = e^{-j2p\mathbf{k}^T \mathbf{t}} \quad (4)$$

Actually, the main drawback of this approach is represented by the accuracy of the rototranslation estimate. Rotations less than 1-2 degrees cannot be detected and the accuracy of the translation depends upon the rotation estimate, since translation is computed after the rotation, as mentioned before. On the other hand, though the ICP is a global registration method, it belongs to greedy class of alignement algorithms, meaning that it tends to converge towards a local minimum rather to the global one. In order to align correctly two range views a global minimum has to searched for. Therefore the ICP can be succesfully used only if a good initial approximate of the alignement is known. To this aim we developed a subsystem for registration refinement, where these algorithms are in sequence applied: firstly the Frequency Domain and then the ICP. In this way we are able to solve for the two aforementioned issues: the low accuracy of the frequency technique (improved by the ICP) and the need for an initial good estimate (provided by the FD).

Once the final transformation, which aligns in a least square sense the overlapping area between the two views A and B, has been estimated, it is applied to both the whole views in order to register them. The framework of the 2nd subsystem is depicted in figure 2.

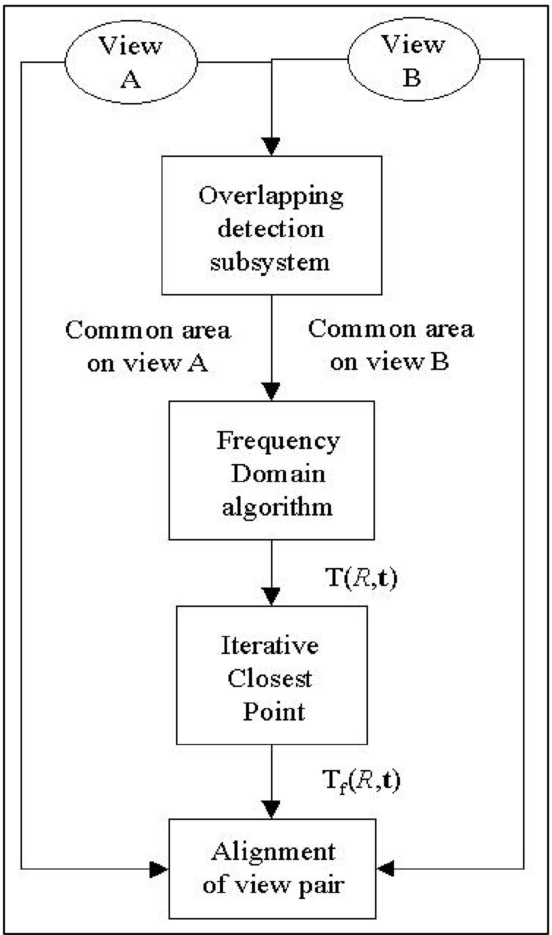

Figure 2: Framework of registration refinement subsystem

4. FUTURE DEVELOPMENTS

The registration system described so far is able to align only range view pairs of a real object. Anyway, in order to build a complete 3D model of a whole object, several views have to be registered together. Therefore, for the next future effort will be spent to develop a full automatic registration system, which is able to work as a black box. More in detail, in a first stage such a system should take as input all the range images acquired by the laser scanner, and register them two by two. At this step there is no need for the user to order the sequence of views according to their adiacency: the system istelf will automatic detect this geometrical relationship during pairwise alignement. It's clear that in this way, the registration process could be greatly enhanced, since the first part of the modeling process would be done by a machine: the only limit will depend upon the complexity of the shape of the object, upon the amount of 3D data acquired and the computing preformance. In the second stage of such automatic system, once the range views were correctly aligned, a global registration algorithm, like ICP or one of its variants, could be applied in order to get the final registration of the whole set of range views. A possible framework of proposed system is presented in figure 3.

Another interesting topic for future developments, deals with the optimization of memory requirements, and therefore the speeding-up of the registration procedure. This represents still today an issue, overall when facing with large datasets, comprising of millions of 3D points, which are acquirable by modern terrestrial laser scanners. In this field further investigations could be undertaken in order to define a coding of the object shape, still based on spin-images but reducing the redundancy, that is always present in terms of high degree of similarity between spin-images of adjacent surface points. Suitable form of spin-image compression has to be investigated, which allows to effectively compare compressed spin-images. Moreover, even the mesh resolution plays a major role in memory occupancy, as it defines the quantization step for spin-image generation. In this case a solution could be the development of pyramidal matching procedure, that according to a coarse-to-fine approach is able to determine the correspondences using meshes of different resolutions, increasing step by step the level of detail of the mesh.

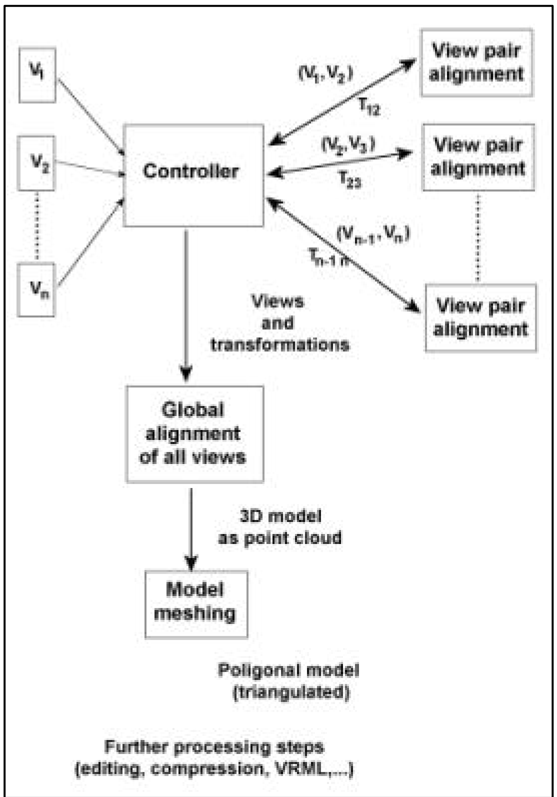

Figure 3: Framework of global alignment system

5. CONCLUSIONS

In this paper the framework of an automatic registration system of range data has been presented. The main goal of this project is aimed to reduce as more as possible the time spent for the alignement process, which today still requires an initial approximate estimate to be manually defined by the user. Basically the procedure is based on the application of spin-image concept to triangulated meshes of two views and the refinement of a first rough transformation estimate by the sequential application of two global alignement algorithms, i.e. the Frequency Domain and the ICP. Spin-images provide a new kind of object shape encoding, where the global properties of the object are retained despite of the specific position of surface points. Then, through spin-images surface matching can be turned into the recognition of the most similar images of two image sets, a well studied problem for which a number of techniques are available. This approach yields an approximate transformation between overlapping area of an image pair, that has to be refined in order to compute a correct registration between full views. To this aim two global alignement algorithms were applied, the Frequency Domain and ICP. The first one allows to refine the estimate through separate computation of rotation and translation using Fourier transform, and in the same time it provides a good initial estimate for subsequent application of the ICP. In turn, this algorithm provides the optimal transformation not only between the overlapping areas but also between the whole data set of a view pair. Further improvements has to be carried out, in terms of optimization of memory occupancy and speeding-up of the whole process, overall when facing with data sets composed by million of points, as the ones acquired by todays available terrestrial laser scanner.

Some interesting results of the application of the registration system are depicted in figures 4-8. More in detail, figures 4a and 4b show two adjacent scans of Murer's statue acquired with our close-range laser scanner BIRIS, while figure 4 shows the result of the first approximate alignment with the detected common area between such scans. Then, in figure 6 a scan of the foreground of an ancient castle located in Cento (Fe) is depicted. Here range data were collected with the Cyrax 2500 laser scanner. The spatial resolution was set to 3 cm. Figures 7 and 8 show the results of the alignment performed with both subsystems. Actually, given the large amount of the points, only a subset of them was used in both scans to estimate the common area and to perform the image registration.

ACKNOWLEDGMENTS

This work was developed with the project "Application in the survey, store and management of environmental and cultural resources of GNSS/INS positioning and satellite, aerial, terrestrial photographic and laser scanning data, transmitted by DARC, GSM/GLOBAL STAR, INTERNET methods", partly financed by MURST (Italian Ministry of University and Research) in 2002 as project of relevant National interest. National coordinator: Giorgio Manzoni, head of the Research unit Antonio Vettore.

REFERENCES

Besl P. J., McKay N. D., 1992. A Method for Registration of 3D Shapes. *IEEE Trans. on PAMI*, 14(2): 239-256.

Bologna R., Guarnieri A., Minchilli M., Vettore A., 2002. Automatic registration of 3D views, *Proceedings of the ISPRS Comm. V Symposium "Close Range Imaging-Long-Range Vision"*, 2-6 September, Corfù, Greece.

El-Hakim S. F., 2001. 3D Modeling of Complex Environments. Videometrics and Optical Methods for 3D Shape measurement, *Proceedings of SPIE* , vol 4309.

Guarnieri A., Marton F., Vettore A., 1999. 3D Modeling of Real Artistic Objects with Limited Computers Resources. *Proc. of XVIII CIPA Symposium on Architectural and Archaeological Photogrammetry*, Recife, Brasil, 3-8 October.

Guidi G., Beraldin J.-A., Ciofi S., Atzeni C. Fusion of range camera and photogrammetry: a systematic procedure for improving 3D models metric accuracy. Accepted on march 2003 for publication on the *IEEE Trans. on Systems, Man, and Cybernetics*.

Horn B. K. P., 1987. Close-form solution of absolute orientation using unit quaternions. *Journal of Optical Society of America*. A, 4(4):629-642.

Italy-Canada 2001 Workshop on "3D Digital Imaging and Modeling Applications of: Heritage, Industry, Medicine & Land", Padua, Italy , 3-4 April.

Johnson A. E.,1997. Spin-Images: A Representation for 3D Surface Matching". Ph. D. Thesis, CMU-RI-TR-97-47, Carnegie Mellon University, USA.

Lucchese L., Cortelazzo G. M., Vettore A., 1997. Estimating 3D Rototranslations from Range Data by a Frequency Domain Technique. *Proceedings of Optical 3D Measurement Techniques IV*, Zuerich, Switzerland.

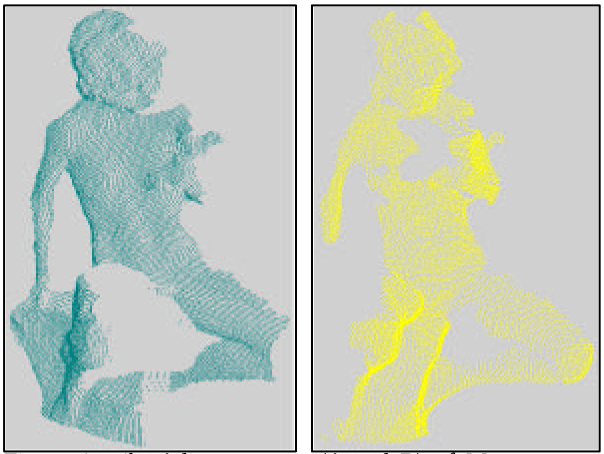

Figure 4a, 4b: Adjacent scans (A and B) of Murer statue acquired by BIRIS close-range laser scanner

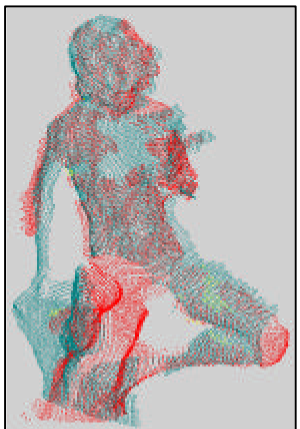

Figure 5: Front view of Murer, showing the common area between mesh A and B as detected by the first subsystem.

Figure 6: Front view of scan1 of Cento's castle acquired by Cyrax 2500

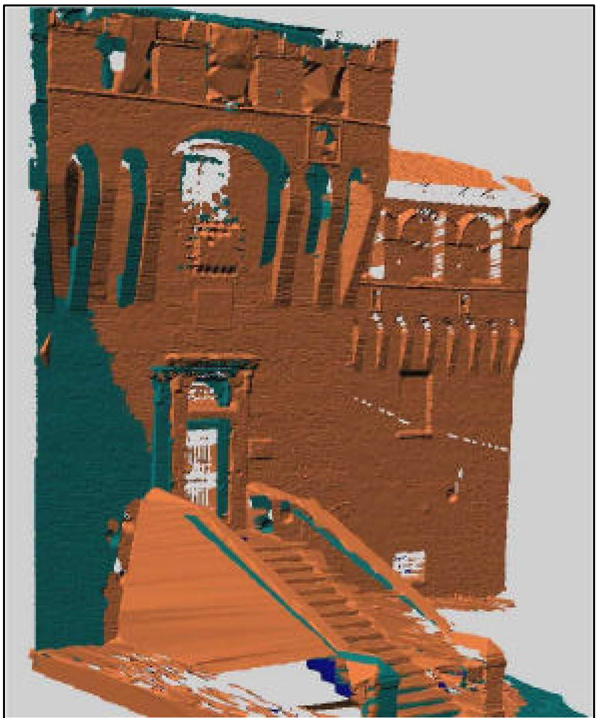

Figure 7: Result of the pre-alignment of the two scans of the castle.

Figure 8: Result of the global (refined) alignment with the detected overlapping area (in red).